Free Energy Calculations in Globular Proteins: Methods to Reduce Errors

ALFREDO DI NOLA,* AXEL T. BRÜNGER

Howard Hughes Medical Institute and Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06511

Received 14 May 1997; accepted 28 February 1998

ABSTRACT: The calculation of free energies by computer simulation represents one of the most promising areas in molecular modeling. While the computational methods developed so far give reliable results for liquids or solutions, they are not satisfactory for globular proteins. The reproducibility of the data is poor due to several sources of error. The most important are due to the magnitude of the molecule's phase space, to the long relaxation time of the system, and to the singularity occurring when creating or annihilating atoms. In a previous study Simonson and Brünger reported the free energy differences calculated for three successive mutations in the ribonuclease-S system and revealed several sources of error. These errors were reanalyzed and the performance of several methods studied in order to reduce them. Different approaches of mutating the Hamiltonian are compared using the method proposed by Resat and Mezei and a modification of the method proposed by Cross. Procedures are also proposed to reduce the effects of the long relaxation time of the molecule, to bias the simulation toward the experimental structure, and to reduce large free energy derivative fluctuations. All these methods give reliable results when the mutation is carried out in a peptide in solution. When the mutation is carried out in a globular protein, the sources of errors are reduced but not eliminated. Although the investigated procedures and methods increase the reliability of free energy calculation, further improvements will be required. J Comput Chem 19: 1229–1240, 1998

Keywords: free energy calculations; molecular dynamics; ribonuclease A

* *Permanent address:* Università degli Studi di Roma "La Sapienza," Dipartimento di Chimica, Ple. A. Moro 5 00185 Roma, Italy

Correspondence to: A. Di Nola

Contract/grant sponsor: National Institute of Health; contract/grant number GM-39546-09

Contract/grant sponsor: CNR-Italy

CCC 0192-8651 / 98 / 111229-12

Introduction

The calculation of free energies by computer simulation represents one of the most promising areas in molecular modeling. However, in contrast to energetic and structural properties, free energies require in principle an infinitely long simulation, because the whole phase space available to the system must be sampled.[1] In addition, the statistical mechanical free energy relationships are based on the assumption that the system is in equilibrium. This implies that proper equilibration and sampling must be performed.[2] These assumptions make the calculation of free energy in globular proteins[3,4] very difficult to perform accurately in contrast to simple liquids. In practice the most severe problems are related to the magnitude of the phase space and to the long relaxation time of the protein compared to the simulation time. Pearlman and Kollman[5] showed that very long simulations (> 2 ns) are required for the mutation of simple solutes in water. Recent results obtained by the essential dynamics method[6,7] showed that the region of the configurational space accessible to a globular protein in solution is much wider than the region sampled by a classical molecular dynamics (MD) simulation. This implies that by increasing the length of the simulations more and more conformations are sampled and the phase space is not adequately sampled.

An additional technical problem is due to the so-called *origin catastrophe* (i.e., the singularity occurring in MD simulations when creating or annihilating atoms). For example, in the thermodynamic integration (TI) method, the mutation of a molecule A into a molecule B is obtained by progressively mutating the Hamiltonian along a chosen pathway and the free energy difference is calculated according to the formula

$$\Delta G = \int_0^1 \langle \frac{\delta H(\lambda)}{\delta \lambda} \rangle \, d\lambda. \qquad (1)$$

If the dependence of the Hamiltonian on λ in (1) is linear, TI leads to an improper integral at the endpoint(s) (i.e., to a definite integral with a singular integrand). For a functional form $1/r^e$ in a space with dimensionality d, it is known that the integrand remains finite everywhere if $H(\lambda) \propto \lambda^k$ with $k \geq d/e$. For reviews, see refs. 2 and 8–11.

Bovine pancreatic ribonuclease A (RNase-A) is an ideal model system for developing free energy methodologies. It can be cleaved between residues 20 and 21, producing the S-peptide and S-protein.[12] The two fragments can be reconstituted to form an enzymatically active complex. Only the first 15 residues of the S-peptide contribute to structure and binding, because residues 16–20 are disordered and a 15-residue S-peptide forms a complex almost identical to the 20-residue peptide. Connelly et al.[13] and Varadarajan et al.[14] designed a series of hydrophobic mutations at position 13 represented by a methionine in the native protein and measured the free energies, enthalpies, and entropies of binding. The crystal structures of the native and the mutants are also available.[15,16]

In a previous article Simonson and Brünger[17] reported the results of the calculation of the free energy differences for the binding of the S-peptide to the ribonuclease S-protein among different mutants of Met13. They showed that one of the major sources of errors is due to the different minima visited during the simulation and suggested that one should bias the simulation toward the known experimental structures. In addition, they suggested carefully designing the mutation pathway in order to reduce the energy buildup in the barrier passage.

Our goal was to reanalyze the previously discussed sources of error and to suggest methods to solve them. In particular, we addressed the following questions: What is the best dependence of the Hamiltonian on the coupling parameter λ? Is it possible to reduce the effect of the long relaxation time of a globular protein? Is it possible to reduce the free energy changes during the simulation? Is it possible to bias the simulation toward the desired (experimental) structure?

Many attempts[18,19] have been made in recent years to describe the appropriate dependence of the Hamiltonian on the coupling parameter λ. They were mainly focused on obtaining a smooth and monotonic varying Hamiltonian without singularity. Although the results of the free energy calculations in principle are independent of the path employed, they actually strongly depend on it due to numerical problems. In the present study we used and compared two different methods: the one proposed by Resat and Mezei[18] and a method similar to the one proposed by Cross.[20]

Due to the long relaxation time of proteins, it is very important to perturb the system as little as possible: large free energy changes during the sim-

ulation require very long simulation times and make the protein sample a large region of the conformational space. To reduce these effects we propose a new pathway for annihilating or creating atoms using a Hamiltonian dependent on two parameters λ_1 and λ_2. We also propose a procedure based on the insertion, in each window, of a restrained equilibration step, consisting of a simulation at high temperature with dihedral angles restrained. This procedure has two advantages: it allows a fast relaxation of the structure and biases the simulation toward the experimental structure. The RNase-S S-peptide Ile → Val and Ile → Leu mutations are used as test cases. These mutations were chosen because they involve the creation and annihilation of one and two atoms, respectively, and are particularly suitable for a methodological study.

Materials and Methods

Our simulations involve the interconversion of residue 13 of the S-peptide from isoleucine to valine and to leucine and the inverse steps. To simulate the mutation of a molecule A into a molecule B, the Hamiltonian is progressively mapped from $H(A)$ to $H(B)$ along some chosen pathway. We set $H = H(\lambda)$, where λ is a coupling parameter that varies from 0 to 1, $H(0) = H(A)$, and $H(1) = H(B)$. We used the thermodynamic integration approach in which the free energy difference between states A and B is estimated by eq. (1). In practice, simulations are done at a few discrete points λ_i along the chosen pathway and the integral is calculated by interpolation. The choice of the pathway is practically very important, and it will be discussed in detail below.

To calculate the $\Delta\Delta G$ of binding of a mutant S-peptide to the S-protein, compared to the native peptide, we used the thermodynamic cycle[21] depicted in Figure 1. Simulation of the peptide in solution yields ΔG(Pep); simulation of the complex in solution yields ΔG(Com). $\Delta\Delta G$ can thus be obtained by

$$\begin{aligned}\Delta\Delta G &= \Delta G'(\text{binding}) - \Delta G(\text{binding}) \\ &= \Delta G(\text{Com}) - \Delta G(\text{Pep}). \end{aligned} \quad (2)$$

Following the work of Simonson and Brünger,[17] an α-helical conformation for the native and mutant S-peptide in solution was used, which was derived from the corresponding crystal structures[15,16] of the complexes. Harmonic restraints to keep the peptide in the α-helical conformation were introduced. The experimental free energy values were taken from Varadarajan et al.[14] Histidines were protonated as described previously.[17]

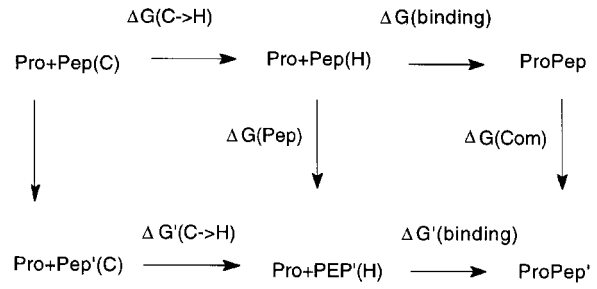

FIGURE 1. Thermodynamic cycle describing S-petptide–S-protein binding. The binding is represented as a two-step process, and the peptide goes from its unfolded conformation in solution, Pep(C), to the α-helical conformation it assumes in the complex, Pep(H), and then bound to the protein.

ENERGY PARAMETERS

The OPLS/AMBER[22,23] force field was used to model the protein, and the TIP3P model was used for water.[24] Polar hydrogens were explicitly included in the model, and hydrogen bonding was modeled as a purely electrostatic interaction. Electrostatic interactions were truncated on a group basis at a distance of 9 Å. We used a dielectric constant of 1.

STOCHASTIC BOUNDARY

MD simulations were carried out with the stochastic boundary method.[25–27] The radius of the spherical region was slightly increased compared to the work by Simonson and Brünger.[17] The radius was set to 14 Å and the sphere centered on the β-carbon of the residue 13. The sphere was filled with water molecules that produced a total of 186 and 311 water molecules in the peptide–protein complex and in the S-peptide, respectively. For each mutation, the crystal structures of the two mutants were superimposed. All protein atoms not included in the spherical region and whose positions differed by less than 0.2 Å were kept fixed if the backbone atoms of both residues coincided. All the remaining atoms were unrestrained. With this procedure about half of the protein was contained in the spherical region. For the S-peptide in solution a few atoms of Lys 1 were kept fixed.

The Verlet algorithm was used to integrate the equations of motion,[28] and a weak coupling to a thermal bath was used to keep the temperature at the desired value.[29] The SHAKE algorithm[30] was used to fix the bond lengths of solute and solvent molecules. A time step of 2 fs was used to integrate the equations of motion, and artificial masses of 10 amu for the solute hydrogens were used.

A preliminary equilibration of the structure in water was performed as follows: in the first step the solute was fixed and the water molecules were relaxed for 10–20 ps at room temperature. Subsequently the system was quenched to 10 K and the constraints on the solute were removed. The temperature of the heat bath was slowly raised and a further 90 ps of equilibration at room temperature were simulated. The final resulting structure was then taken as the starting point of the simulation. All simulations were run with the program X-PLOR.[31]

RANDOM STATISTICAL ERRORS

The random statistical error[32–34] of a free energy difference was calculated from the standard error $\sigma(\bar{a})$ of a time average $\bar{a}$ that depends on the size n of the sample and on the correlations within the sample, according to the formula

$$\sigma(\bar{a}) = \frac{\sigma(a)^2}{n} + \frac{2}{n}\sum_{j=1}^{n-1}\left(1 - \frac{j}{n}\right)K(j), \quad (3)$$

where $\sigma(\bar{a})$ is the standard deviation of the time series $\{a_i\}$ and $K(j)$ is the time-correlation function $K(j) = \overline{a_i a_{i+j}}$, which can be computed from the trajectory.

THERMODYNAMIC INTEGRATION—CHOICE OF PATHWAY: $H(\lambda)$

It is well known that the introduction (or removal) of an atom into a system causes a rapid variation of the free energy at the beginning (or end) of the MD simulation.[8] It is caused by the occurrence of a close contact between two atoms. To avoid this problem several methods have been proposed.[2, 5, 17, 20, 35–38] A simple idea to avoid the problem is shown in Figure 2, which consists of "growing in" an atom bonded to an already existing atom of the molecule. Because the growing atom is partially shielded by its neighbors, no singularity can occur. However, if the van der Waals energy is described through a Lennard–Jones function,

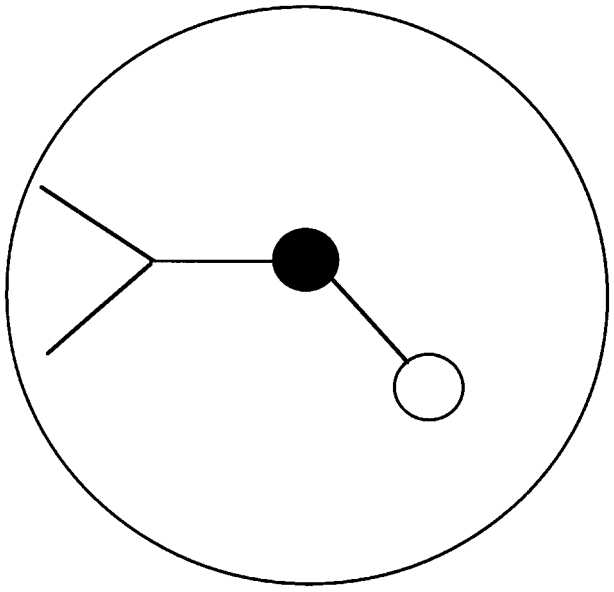

FIGURE 2. Shielding sphere produced by the atom represented by filled circles.

$$V_{ij}(\lambda) = 4\lambda\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right], \quad (4a)$$

the fractions in parentheses can assume very large values for $r_{ij} \ll \sigma_{ij}$ during the "growing" process ($\lambda \to 1$), thus causing numerical problems when free energy averages are computed.

In the method proposed by Cross[20] ε_{ij} and σ_{ij} are both linearly dependent on λ so that

$$V_{ij}(\lambda) = 4\lambda\varepsilon_{ij}\left[\left(\frac{\lambda\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\lambda\sigma_{ij}}{r_{ij}}\right)^{6}\right]. \quad (4b)$$

In computer programs such as X-PLOR, technical complications arise due to the dependence of ε_{ij} and σ_{ij} on individual ε and σ values of atoms i and j:

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \cdot \varepsilon_j} \quad \text{and} \quad \sigma_{ij} = \frac{\sigma_i + \sigma_j}{2}.$$

We have thus modified Cross' method as follows:

$$V_{ij}(\lambda) = 4\varepsilon_{ij}(\lambda)\left[\left(\frac{\sigma_{ij}(\lambda)}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}(\lambda)}{r_{ij}}\right)^{6}\right], \quad (4c)$$

with

$$\varepsilon_{ij}(\lambda) = \lambda\sqrt{\varepsilon_i \cdot \varepsilon_j} \quad \text{and}$$

$$\sigma_{ij}(\lambda) = \frac{(2\lambda - 1)\sigma_i + \sigma_j}{2}.$$

In particular, we obtain $\sigma_{ij}(0) = (-\sigma_i + \sigma_j)/2$ and $\sigma_{ij}(1) = (\sigma_i + \sigma_j)/2$ so that σ_{ij} is small for $\lambda \to 0$. The shielding by the atom bonded to atom i avoids numerical convergence problems. As λ approaches zero, $r_{ij} \neq 0$ and $\sigma_{ij} \approx 0$. Hereafter this method will be referred to as the nonlinear (NL) method.

In the method proposed by Resat and Mezei,[18] the interaction potential between two nonbonded atoms is made dependent on different powers of the parameter λ.

$$V_{ij}(\lambda) = \lambda^4 \frac{A_{ij}}{r^{12}} - \lambda^3 \frac{B_{ij}}{r^6} + \lambda^2 \frac{q_i q_j}{r}. \qquad (4d)$$

This method will be referred to as the polynomial path (PP) method. The main difference between the PP and NL methods is that the variation of the Hamiltonian is smoother in the former method. It should be noted that a common choice of the pathway in thermodynamic integration consists of the dependence of the Hamiltonian on a power k of the parameter λ with $k \geq 4$. The choice of $k = 4$ produces a Hamiltonian that varies in a somewhat intermediate way compared to the NL and PP methods.

CRYSTAL STRUCTURE RESTRAINTS

Free energy perturbation calculations are usually performed in three stages at each value of the parameter λ: energy minimization, equilibration, and data acquisition. Simonson and Brünger[17] used this procedure and found that the final conformation of the mutated residue can be different from that found in the mutant crystal structure. We have therefore modified the procedure as follows: a new step is inserted after the energy minimization stage in which the dihedral angles around the mutation site are restrained to those of the crystal structure. Thus, the stages of the modified protocol are as follows: minimization, restrained relaxation, equilibration, and acquisition. This procedure is valid if the experimentally observed rotamers are populated, even though not preferred, at each window, which is probably a reasonable assumption for the present system. During the subsequent equilibration and acquisition stages, it is possible that rotamers different from the experimental ones are populated. It was indeed found that a few dihedral transitions occurred at the beginning of the growth of one of the atoms. A further advantage of this protocol is that the temperature of the relaxation step can be increased to allow better relaxation of the molecule.

The restraining potential is removed in the subsequent steps so that it is not necessary to compensate for it in the free energy calculation. All the reported mutations were performed with a 30-ps simulation for each window (10-ps restrained relaxation + 10-ps equilibration + 10-ps acquisition). Tests performed with twice the simulation time did not produce significantly different results.

The starting structure was obtained as described in the Materials and Methods Section. The backbone root mean square (RMS) deviation between the starting M13ILE structure in water after equilibration and crystal structure is less than 1 Å, with the exception of three regions: the N-terminal regions of the S-peptide and S-protein, and the region between residues 48 and 52 of the S-protein.

Results and Discussion

DUMMY MUTATION: (Ile → Ile)

The performance of different protocols for the free energy calculations was first studied for the dummy mutation (Ile → Ile) shown in Figure 3: atom CDA was created and atom CDB removed. In the first calculation we applied the NL method [eq. 4(c)] using 23 windows and

$$\lambda_{\mathrm{CDA}} = \lambda_{\mathrm{in}} = t,$$

$$\lambda_{\mathrm{CDB}} = \lambda_{\mathrm{out}} = 1 - t \quad \text{for} \quad 0 \leq t \leq 1. \qquad (5)$$

At $t = 0$, $\lambda_{\mathrm{CDA}} = 0$ and $\lambda_{\mathrm{CDB}} = 1$ so that atom A is a ''ghost'' atom and only atom B is present. The reverse situation occurs for $t = 1$. The path in the λ_{in}, λ_{out} plane is shown in Figure 4 (curve a).

The free energy derivatives obtained by using a temperature of the restraining step simulation of 300 K, are reported in Figure 5, (curve a). It should be noted that λ in Figure 5 represents the pathlength of the corresponding curve of Figure 4. As a consequence, its final value is greater than 1. The expected value of ΔG is 0 kcal/mol, and the calculated free energy difference obtained is ΔG = 2.8 (SD 0.16) kcal/mol. The large free energy difference has to be attributed to the long relaxation

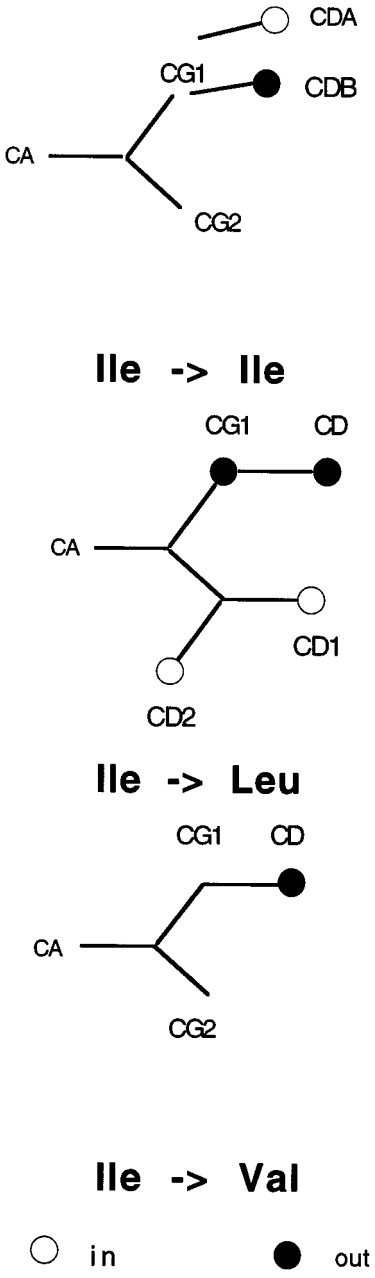

FIGURE 3. Schematic view of the hybrid amino acids involved in the various mutations. Changes due to differences in geometry and partial charges are not indicated.

time of the protein and to a large perturbation of the system.

The mutation was repeated by raising the temperature of the restrained relaxation step to $T = 330$ K (Fig. 5, curve a′). The free energy difference obtained was $\Delta G = -0.17$ (SD 0.14) kcal/mol, which coincides with the expected value within 2 SD. The curves reported in Figure 5 show that the

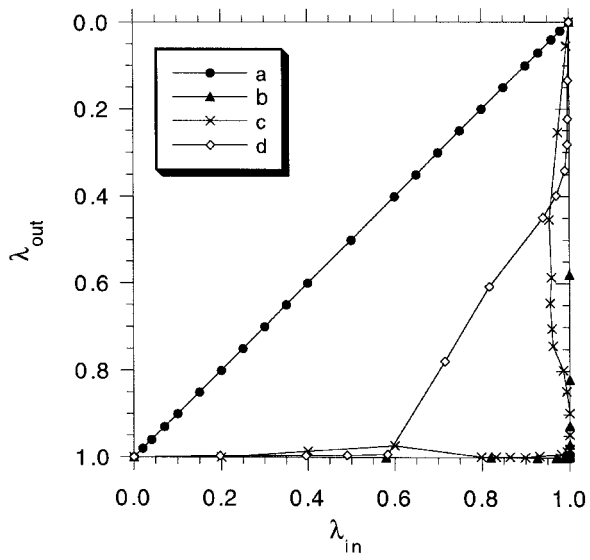

FIGURE 4. Path of the different M13Ile to M13Ile mutations in the λ_{in}, λ_{out} plane. (Curve a) NL method with $\lambda_{in} = t$ and $\lambda_{out} = 1 - t$; (curve b) NL method using eq. (6) with $r_1 = r_2 = 0.07$; (curve c) NL + SPM methods with two independent λ_{in} and λ_{out}; (curve d) PP + SPM methods with two independent λ_{in} and λ_{out}.

mutation pathway consists of two well-separated stages: a first stage dominated by the influence of the removed atom and a second stage dominated by the created atom. The free energy derivative assumes very large values corresponding to a large perturbation of the system.

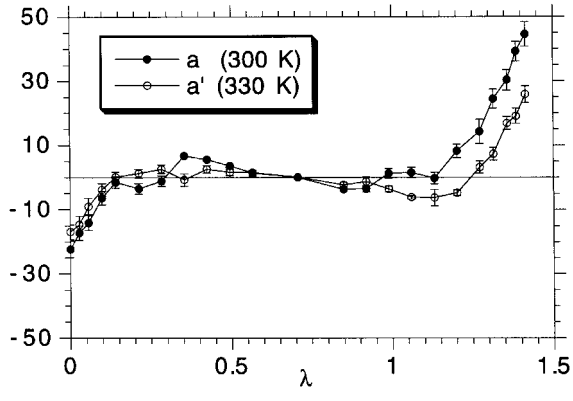

FIGURE 5. Free energy derivative with respect to λ for the M13Ile ↦ M13Ile dummy mutation in the complex, corresponding to curve a of Figure 4. (●, curve a) temperature of the restrained relaxation step $T = 300$ K; (○, curve a′) temperature of the restrained relaxation step $T = 330$ K. λ represents the length of the corresponding path in Figure 4.

To reduce the large free energy fluctuations we used a different t dependence of the parameters λ_{in} and λ_{out} (i.e., a different path in the λ_{in}, λ_{out} plane). Preliminary tests showed that a strong reduction of the free energy fluctuation can be obtained by first creating atom CDA followed by removal of atom CDB. Such a path can be obtained by the following dependence of λ_{in} and λ_{out} on t:

$$\lambda_{in} = c(1 - e^{(t/r_1)}) \quad \text{and}$$
$$\lambda_{out} = 1 - d(e^{(t/r_2)} - 1). \qquad (6)$$

The r_1 and r_2 were adjustable parameters and c and d were chosen such that at $t = 1$, $\lambda_{in} = 1$ and $\lambda_{out} = 0$, respectively. (Note that a linear dependence of λ_{in} and λ_{out} on t would correspond to $r_1 = r_2 = \infty$.)

The path obtained for $r_1 = r_2 = 0.07$, using 16 windows, is reported in Figure 4 (curve b). The path consists of an initial, almost complete, creation of atom A followed by the removal of atom B. In Figure 6 (curve b) the free energy derivative is shown versus the path length. The comparison of this curve to the ones reported in Figure 5 shows that the maximum value of the free energy derivative is strongly reduced. The free energy difference obtained was $\Delta G = 0.28$ (SD 0.18) kcal/mol, which is close to the experimental value. In this case different temperatures of the restrained step gave comparable results, so it was not necessary to increase the temperature of the restrained equilibration step. These results show that a properly chosen path can improve the reliability of the results.

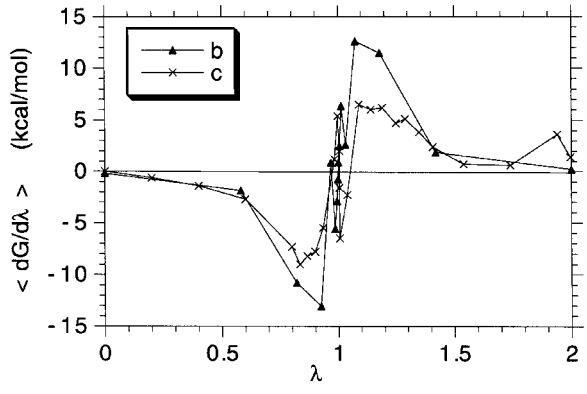

FIGURE 6. Free energy derivative with respect to λ for the M13Ile → M13Ile dummy mutation in the complex, corresponding to curves (▲) b and (×) c of Figure 4. λ represents the length of the corresponding path in Figure 4.

To further reduce the free energy fluctuations, we completely decoupled the two parameters λ_{in} and λ_{out} using a slightly different procedure. The goal of this procedure was to follow the smoothest path in the free energy landscape. The introduction of two parameters has been used previously for decoupling the van der Waals and electrostatic terms and/or for distinguishing among different molecules.[39] At each window the derivative of the Hamiltonian with respect to both parameters is calculated and the new values of the parameters are chosen according to the following two criteria: the distance between the final point ($\lambda_{in} = 1$, $\lambda_{out} = 0$) and the current point must decrease; the direction and the length of the new step are determined in such a way that the estimated contribution of the new step to the free energy must be lower than a maximum chosen value. In this way the number of windows is a dynamic parameter. This method will be referred to as the smooth path method (SPM).

The results obtained with the NL method and SPM are reported in Figures 4 (curve c) and 6 (curve c). The number of windows was 26. Figure 6 shows that a further reduction of the maximum value of the free energy derivative was achieved. The free energy difference obtained was $\Delta G = -0.28$ (SD 0.13) kcal/mol. We now compare the results obtained with the NL method for the three different paths in the λ_{in}, λ_{out} plane: curves a and a′ of Figure 5 corresponding to path a of Figure 4; curve b of Figure 6 corresponding to path b of Figure 4 [eq. (6)]; curve c of Figure 6 corresponding to path c of Figure 4 obtained with the SPM. The maximum value of the free energy derivative decreases from curve a to c. This decrement corresponds to a smoother path in the free energy landscape. Thus, the SPM is preferable to other methods.

To compare different ways of mutating the Hamiltonian, we performed to same mutation by using the PP method of Resat and Mezei[18] [eq. (4d)]. In the present case the interaction potentials are given by

$$V(\lambda_{in}) = \lambda_{in}^4 \frac{A_{ij}}{r_{ij}^{12}} + \lambda_{in}^3 \frac{B_{ij}}{r_{ij}^6} \quad \text{and}$$
$$V(\lambda_{out}) = \lambda_{out}^4 \frac{A_{ij}}{r_{ij}^{12}} + \lambda_{out}^3 \frac{B_{ij}}{r_{ij}^6},$$

and the SPM method was used. The resulting path is shown in Figure 4 (curve d), and the dependence of the free energy derivative is reported in

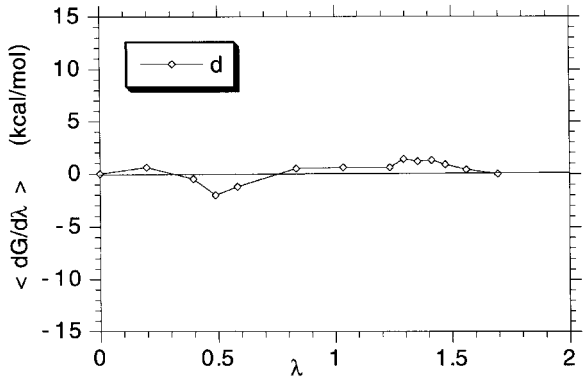

FIGURE 7. Free energy derivative with respect to λ for the M13Ile → M13Ile dummy mutation in the complex, corresponding to (◇) curve d of Figure 4. λ represents the length of the corresponding path in Figure 4.

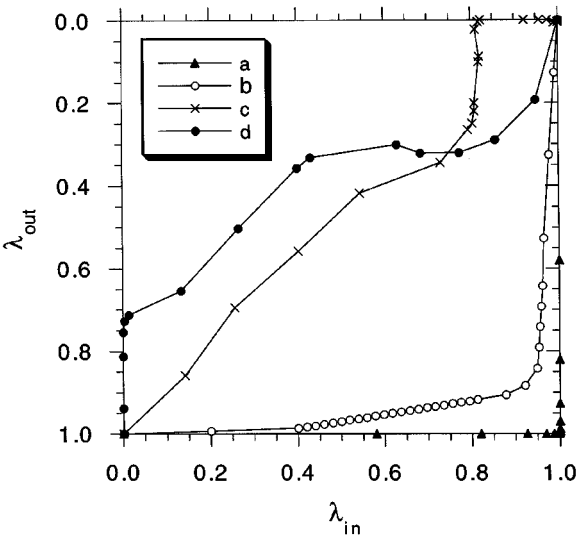

FIGURE 8. Path of the different M13Ile ↔ M13Leu mutations of the peptide in the λ_{in}, λ_{out} plane. (Curve a) NL method using eq. (6) with $r_1 = r_2 = 0.07$; (curve b) NL + SPM methods with two independent parameters, λ_{in} and λ_{out}, for the forward mutation; (curves c, d) PP + SPM methods with two independent λ_{in} and λ_{out} for forward and backward mutations, respectively.

Figure 7. Figure 7 shows that the maximum value of the free energy derivative is further reduced and that the curve is smoother than the one obtained by the NL method. The calculated value of the free energy difference was $\Delta G = 0.31$ (SD 0.15) kcal/mol, which is close to the expected value within ~ 2 SDs. In this case 14 windows were necessary compared to the 26 windows for the NL + SPM method. It can be concluded that the combined use of SPM for the path in the λ_{in}, λ_{out} plane and the PP method for the dependence of the Hamiltonian on the parameters produces the smoothest path in the free energy landscape.

Ile ↔ Leu MUTATION

In this section we report the results obtained for the (Ile ↔ Leu) mutation that involves creation of two atoms and simultaneous removal of two atoms, as illustrated in Figure 3. The backbone RMS deviation between the mutant crystal structures is less than 0.5 Å. Different paths and different ways of mutating the Hamiltonian of the created and removed atoms were tested using both the NL and PP methods. It should be pointed out that for all mutations the final structures after thermodynamic integration have the correct rotamers around the mutation site.

The paths in the λ_{in}, λ_{out} plane for the mutation of the peptide are shown in Figure 8. Curves a correspond to the NL method using eq. (6) for the dependence of λ_{in} and λ_{out} on t for the forward and backward motion (16 windows), curve b the NL + SPM methods for the forward mutation (35 windows), and curves c and d the PP + SPM methods for forward and backward mutations, respectively (16 windows). The corresponding free energy derivatives are reported in Figure 9a and b for the NL and PP methods, respectively.

The calculated free energy differences for each simulation are reported in Table I. As in the case of the Ile ↦ Ile mutation, the PP method produces lower values for the free energy derivative and the curves are more monotonic and smoother than the NL method. The calculation of the free energy difference in the protein was performed following the paths indicated in Figure 10, curve a corresponding to the NL method using eq. (6) for forward and backward mutations (16 windows), curve b the NL + SPM methods for the forward mutation (33 windows), and curves c–f the PP + SPM methods using different paths and number of windows (21, 21, 21, and 13). The values of the free energy differences are reported in Table II. In contrast to what is observed for the peptide, the results for the protein are strongly dependent on the path and on the method used. We attribute this result to the complexity of the mutation (involving four atoms) that perturbs the protein structure and causes problems related to the long relaxation time and to the magnitude of the conformational space sampled.

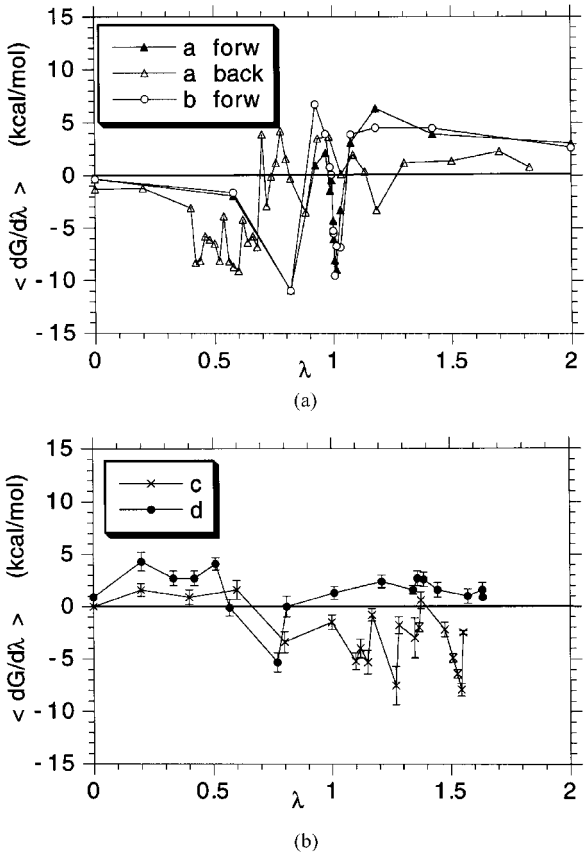

FIGURE 9. Free energy derivative with respect to λ for the M13Ile ↔ M13Leu mutation in the peptide. (a) (▲) Forward and (△) backward mutations corresponding to curve a of Figure 8; (○) forward mutation corresponding to curve b of Figure 8. (b) (×) Forward and (●) backward mutation corresponding to curves c and d of Figure 8.

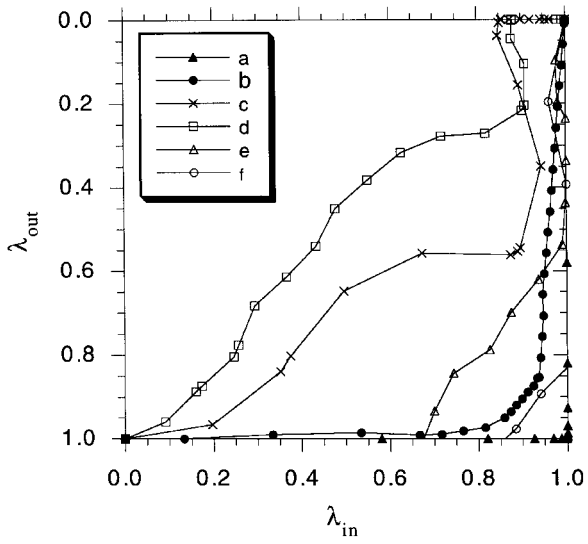

FIGURE 10. Path of the different M13Ile ↔ M13Leu mutations of the complex in the λ_{in}, λ_{out} plane. (Curve a) NL method using eq. (6) with $r_1 = r_2 = 0.07$ (16 windows); (curve b) NL + SPM methods with two independent parameters, λ_{in} and λ_{out}, for the forward mutation (33 windows); (curves c–f) PP + SPM methods using different paths and number of windows of 21, 21, 21, and 13, respectively.

We performed the mutation of the protein in two steps: in the first step atom CD1 is created and atom CD is removed. In the second step atom CD2 is created and atom CG1 is removed (see Fig. 3). Because the PP + SPM method produces lower values of the free energy derivative and less fluctuating curves, these mutations were performed using the PP + SPM method only. The results are reported in Table III. The forward and backward results are comparable within 2–3 SDs. The second step showed very small values of the free energy difference with an average value of −0.18 (0.18). The total free energy difference for the Ile → Leu mutation of the protein is $\langle \Delta G \rangle = -1.65$ (0.22) kcal/mol. The $\Delta\Delta G_{calc}$ value is +0.11 (0.27) kcal/mol which is in close agreement with the experimental value of $\Delta\Delta G_{exp} = +0.2$ kcal/mol.

The results so far reported show that a proper choice of the method, path, and subsequent mutations produce reliable results. However, it seems very difficult to predict free energy differences in globular proteins in the absence of experimental data. In addition, the statistical error of such calculations is large with respect to the experimental value of the measured $\Delta\Delta G$.

ONE ATOM MUTATION Ile ↔ Val

In this section we discuss the results obtained for the Ile13 ↔ Val13 mutation reported in Figure

TABLE I.
Free Energy Differences (kcal / mol) for the Ile ↔ Leu Mutation of the S-Peptide.

Curve	ΔG
a, forward	−1.84 (0.36)
a, backward	+1.39 (0.35)
b, forward	−1.59 (0.33)
c, forward	−1.80 (0.38)
d, backward	+2.19 (0.32)
Average Ile → Leu	−1.76 (0.16)

Standard deviations are reported in parentheses.

TABLE II.
Free Energy Differences (kcal / mol) for Ile ↔ Leu Mutation of S-Protein.

Curve	ΔG
a, forward	−1.91 (0.29)
a, backward	+2.14 (0.31)
b, forward	−1.07 (0.27)
c, forward	−2.02 (0.35)
d, forward	−0.66 (0.30)
e, backward	−0.46 (0.27)
f, backward	+0.25 (0.31)

Standard deviations are reported in parentheses.

3. The deviation between the atomic positions of the backbone atoms of the two mutants in the crystal structure was less than 0.4 Å with the exception of the first residues in the N-terminal regions of the S-peptide and S-protein. In this mutation only one parameter λ was used, because only one atom was mutated. As in the previous simulation, several different mutations were performed for the peptide and the protein. Figure 11a and b shows the dependence of the free energy derivative with respect to the parameter λ for the S-peptide using the NL and PP methods, respectively. The values of the free energy differences are reported in Table IV. As in the Ile ↔ Leu case, the calculations for the peptide produces good results and the PP method products smoother curves than the NL method.

Figure 12a and b shows the dependence of the free energy derivative with respect to the parameter λ for the protein using the NL and PP methods, respectively. The values of the free energy differences are reported in Table V. Again the calculated free energy difference for the protein is less reliable than that of the peptide. The calculated $\Delta\Delta G$ was +0.12 (0.07) kcal/mol compared to the experimental value of −0.1 kcal/mol. The deviation between the initial and final structures gave an RMS deviation for the backbone atoms of less than

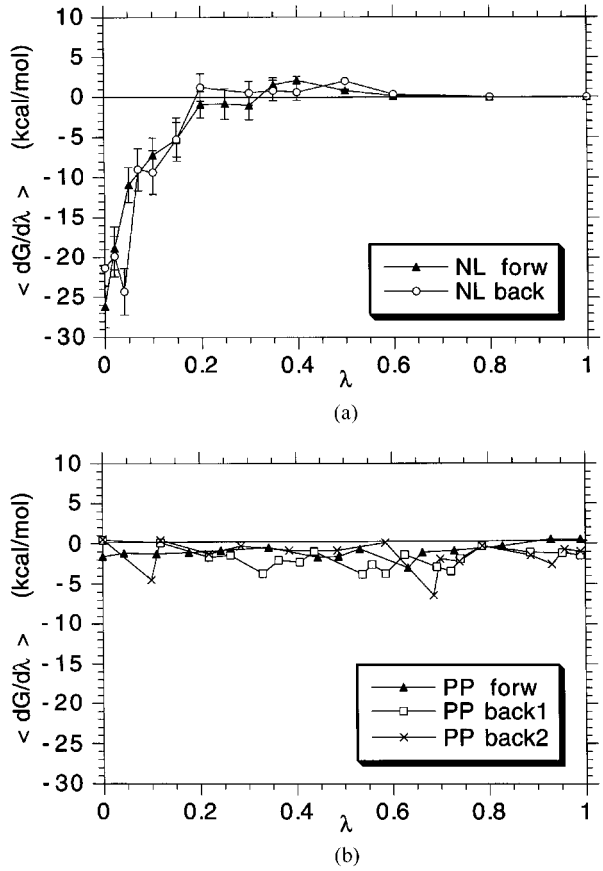

FIGURE 11. (a) Free energy derivative vs. λ for the M13Ile ↔ M13Val mutation in the peptide, obtained with the NL method. (b) Free energy derivative vs. λ for the M13Ile ↔ M13Val mutation in the peptide, obtained with the PP method.

TABLE III.
Free Energy Differences (kcal / mol) for the First Step of Ile ↔ Leu Mutation of S-Protein.

Curve	ΔG
a, forward	−1.74 (0.27)
b, backward	+1.65 (0.29)
c, forward	−1.40 (0.18)
d, backward	+1.07 (0.25)
Average Ile ↔ Leu	−1.47 (0.12)

Standard deviations are reported in parentheses.

TABLE IV.
Free Energy Differences (kcal / mol) for the M13Ile ↔ M13Val Mutation of S-Peptide.

Curve	ΔG
a, NL forward	−1.60 (0.12)
b, NL backward	+1.66 (0.12)
c, PP forward	−1.02 (0.13)
d, PP backward	+1.30 (0.13)
e, PP backward	+1.56 (0.13)
Average Ile → Val	−1.43 (0.06)

Standard deviations are reported in parentheses.

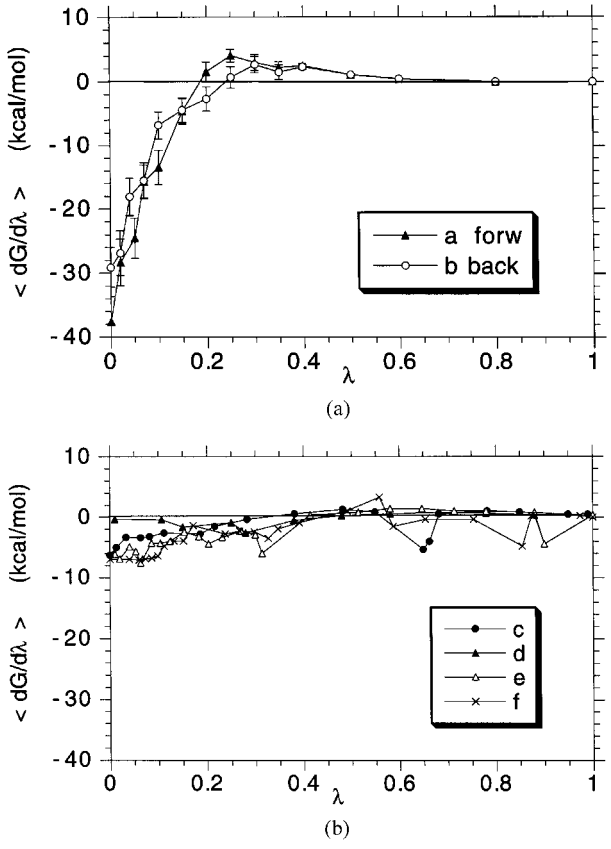

FIGURE 12. (a) Free energy derivative vs. λ for the M13Ile ↔ M13Val mutation in the complex, obtained with the NL method. (b) Free energy derivative vs, λ for the M13Ile ↔ M13Val mutation in the complex, obtained with the PP method.

1.0 Å with the exception of large deviations in the region between residues 49 and 52. The rotamers around the mutation site after thermodynamic integration corresponded to the ones observed in the crystal structure.

TABLE V.
Free Energy Differences (kcal / mol) for the M13Ile ↔ M13Val Mutation of S-Protein.

Curve	ΔG
a, NL forward	−1.97 (0.12)
b, NL backward	+1.78 (0.12)
c, PP forward	−0.62 (0.10)
d, PP backward	+0.43 (0.11)
e, PP backward	+1.27 (0.08)
f, PP forward	−1.64 (0.11)
Average Ile → Val	−1.29 (0.04)

Standard deviations are reported in parentheses.

Conclusions

The goal of this article was to analyze the main sources of error in the free energy calculation of proteins. Two methods of mutating the Hamiltonian were tested: a nonlinear method (NL) and the PP method proposed by Resat and Mezei.[18] The latter method produces smooth free energy derivative curves with relatively small fluctuations. The NL method produces much more fluctuating curves, but the final free energy differences (and standard errors) are comparable to the PP method.

We also used two perturbation parameters that allow a large flexibility in the choice of the path in the λ_1, λ_2 plane and a further reduction of the free energy fluctuations. Finally, we introduced a new step in each window, a dihedral-angle restrained relaxation step, that can be used to increase the temperature and therefore to reduce the effects of the long relaxation time of the protein. This step also allows us to bias the simulation toward the desired (experimental) structure. The restraining potential is removed in the subsequent equilibration and data acquisition steps so that rotamers different from the experimental ones can also be populated.

The use of the methods tested in this study significantly increase the reliability of free energy calculations. However, the results show that the calculated free energy differences are not yet fully satisfactory. The accuracy of the results for the S-peptide is good, but for the S-protein is not yet acceptable. Moreover, the statistical error is too large compared to the experimental data. At present some hypotheses can be made to explain the lower accuracy of the data in the protein with respect to those of the peptide. The main differences between the two systems, from the computational point of view, are that the mutation in the peptide occurs in a residue in close contact with water, while in the protein it occurs in a hydrophobic pocket. The relaxation behavior of the two systems is different. This could imply that longer simulations can improve the results of the protein. However, test calculations show that an increase of the length or of the number of windows does not significantly affect the quality of the results. The second difference is whereas the S-peptide can easily sample the relevant allowed conformational space, the conformational space of the S-protein is much wider and depends on the value of the

coupling parameter(s). Finally, differences between simulation and experimental conditions may be important, such as ionic strength and the pH of the aqueous solution.

Acknowledgments

We acknowledge stimulating discussions with T. Simonson. Support is acknowledged from the National Institute of Health to A.T.B. (GM-39546-09) and CNR-Italy to A.D.N.

References

1. P. M. King, In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Vol. 2, W. F. van Gunsteren, P. K. Wiener, and A. J. Wilkinson, Eds., ESCOM Science Publishers, Leiden, 1993, p. 267.
2. W. F. van Gunsteren, T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, and P. E. Smith, In *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, Vol. 2, W. F. van Gunsteren, P. K. Wiener, and A. J. Wilkinson, Eds., ESCOM Science Publishers, Leiden, 1993, p. 315.
3. H. Lui, A. E. Mark, and W. F. van Gunsteren, *J. Phys. Chem.*, **100**, 9485 (1996).
4. J. L. Miller and P. A. Kollman, *J. Phys. Chem.*, **100**, 8587 (1996).
5. D. Pearlman and P. A. Kollman, *J. Chem. Phys.*, **90**, 2460 (1989).
6. A. Amadei, B. M. Linssen, and H. J. C. Berendsen, *Proteins: Struct., Funct., Genet.*, **17**, 412 (1993).
7. B. L. de Groot, A. Amadei, D. M. F. van Aalten, and H. J. C. Berendsen, *J. Biomol. Struct. Dyn.*, **13**, 741 (1996).
8. D. Beveridge and F. Di Capua, *Annu. Rev. Biophys. Chem.*, **18**, 431 (1989).
9. W. Jorgensen, *Acc. Chem. Res.*, **22**, 184 (1989).
10. W. F. van Gunsteren and P. K. Wiener, Eds., *Computation of Free Energy Biomolecular Systems*, ESCOM Science Publishers, Leiden, 1989.
11. A. Warshel and J. Aqvist, In *Theoretical Biochemistry and Molecular Biophysics*, D. L. Beveridge and R. Lavervy, Eds., Adenine Press, New York, 1989, p. 257.
12. F. M. Richards and H. W. Wyckoff, In *The Enzymes*, Vol. IV, 3rd ed., P. D. Boyer, Ed., Academic Press, New York, 1971, p. 647.
13. P. Connelly, R. Varadarajan, J. Sturtevant, and F. Richards, *Biochemistry*, **29**, 6108 (1990).
14. R. Varadarajan, P. Connelly, J. Sturtevant, and F. Richards, *Biochemistry*, **31**, 1421 (1992).
15. E. Kim, R. Varadarajan, H. Wyckoff, and F. Richards, *Biochemistry*, **31**, 12304 (1992).
16. R. Varadarajan and F. Richards, *Biochemistry*, **31**, 12315 (1992).
17. T. Simonson and A. T. Brünger, *Biochemistry*, **31**, 8661 (1992).
18. H. Resat and M. Mezei, *J. Chem. Phys.*, **99**, 6052 (1993).
19. M. Mezei, *Mol. Simul.*, **10**, 225 (1993).
20. A. J. Cross, *Ann. N.Y. Acad. Sci.*, **482**, 89 (1986).
21. B. Tembe and J. A. McCammon, *Comput. Chem.*, **8**, 281 (1984).
22. W. Jorgensen and J. Tidao−Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).
23. S. Wiener, P. Kollman, D. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. Wiener, *J. Am. Chem. Soc.*, **106**, 765 (1984).
24. W. Jorgensen, J. Chandrasekar, J. Madura, R. Impev, and M. Klein, *J. Chem. Phys.*, **79**, 926 (1983).
25. C. Brooks and M. Karplus, *J. Chem. Phys.*, **79**, 6312 (1984).
26. C. Brooks, A. T. Brünger, and M. Karplus, *Biopolymers*, **24**, 843 (1985).
27. A. Brünger, C. Brooks, and M. Karplus, *Proc. Natl. Acad. Sci. U.S.A.*, **82**, 8458 (1985).
28. L. Verlet, *Phys. Rev.*, **159**, 98 (1967).
29. H. J. C. Berendsen, J. Postma, W. F. van Gunsteren, A. Di Nola, and J. Haak, *J. Chem. Phys.*, **811**, 3684 (1984).
30. J. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.*, **23**, 327 (1977).
31. A. T. Brünger, *X-PLOR Version* 2.1, Yale University, New Haven, CT, 1987.
32. W. Davenport, Jr., *Probability and Random Processes*, McGraw−Hill, New York, 1970.
33. S. Schiferl and D. Wallace, *J. Chem. Phys.*, **83**, 5203 (1983).
34. T. P. Straatsma, H. J. C. Berendsen, and A. Stam, *Mol. Phys.*, **57**, 89 (1986).
35. M. R. Mruzik, F. F. Abraham, D. E. Schreiber, and G. M. Pound, *J. Chem. Phys.*, **64**, 481 (1976).
36. M. Mezei and D. L. Beveridge, *Ann. N.Y. Acad. Sci.*, **482**, 1 (1986).
37. T. Simonson, *Mol. Phys.*, **80**, 441 (1993).
38. P. R. Gerber, A. E. Mark, and W. F. van Gunsteren, *J. Comput.-Aided Mol. Design*, **7**, 305 (1993).
39. X. Kong and C. Brooks, *J. Chem. Phys.*, **105**, 2414 (1996).